

**CLINICAL TRIALS MANAGEMENT SYSTEM AND METHOD**

Inventors:

Michael G. Kahn, M.D., Ph.D.  
Michael Mischke-Reeds

Exp. Mail No.: *E2553343329US*

Attorney Docket No.: FAST-01000US0 WSW  
/wsw/fast/1000.app

# CLINICAL TRIALS MANAGEMENT SYSTEM AND METHOD

## Inventors:

Michael G. Kahn, M.D., Ph.D.  
Michael Mischke-Reeds

## BACKGROUND

### 1. Field of the Invention

The invention relates to the field of medical informatics, and more particularly to a system and method using medical informatics primarily to plan, conduct and analyze clinical trials and their results.

### 2. Description of Related Art

Over the past number of years, the pharmaceutical industry has enjoyed great economic success. The future, however, looks more challenging. During the next few years, products representing a large percentage of gross revenues will come off patent, increasing the industry's dependence upon new drugs. But even with new drugs, with different companies using the same development tools and pursuing similar targets, first-in-category market exclusivity has also fallen dramatically. Thus in order to compete effectively in the future, the pharmaceutical industry needs to increase throughput in clinical development substantially. And this must be done much faster than it has in the past - time to market is often the most important factor driving pharmaceutical profitability.

#### A. **Clinical Trials: the Now and Future Bottleneck**

In U.S. pharmaceutical companies alone, a huge percentage of total annual pharmaceutical research and development funds is spent on human clinical trials. Spending on clinical trials is growing at approximately 15% per year, almost 50% above the industry's sales growth rate. Trials are growing both in number and complexity. For example, the average new

drug submission to the U.S. Food & Drug Administration (FDA) now contains more than double the number of clinical trials, more than triple the number of patients, and a more than a 50% increase in the number of procedures per trial, since the early 1980s.

An analysis of the new drug development process shows a major change in the drivers of time and cost. The discovery process, which formerly dominated time to market, has undergone a revolution due to techniques such as combinatorial chemistry and high-throughput screening. The regulatory phase has been reduced due to FDA reforms and European Union harmonization. In their place, human clinical trials have become the main bottleneck. The time required for clinical trials now approaches 50% of the 15 years or so required for the average new drug to come to market.

## **B. The Trial Process Today**

The conduct of clinical trials has changed remarkably little since trials were first performed in the 1940's. Clinical research remains largely a manual, labor-intensive, paper based process reliant on a cottage industry of physicians in office practices and academic medical centers.

### **1. Initiation**

A typical clinical trial begins with the construction of a clinical protocol, a document which describes how a trial is to be performed, what data elements are to be collected, and what medical conditions need to be reported immediately to the pharmaceutical sponsor and the FDA. The clinical protocol and its author are the ultimate authority on every aspect of the conduct of the clinical trial. This document is the basis for every action performed by multiple players in diverse locations during the entire conduct of the trial. Any deviations from the protocol specifications, no matter how well intentioned, threaten the viability of the data and its usefulness for an FDA submission.

The clinical protocol generally starts with a cut-and-paste word-processor approach by a medical director who rarely has developed more than 1-2 drugs from first clinical trial to final regulatory approval and who cannot reference any historical trials database from within his or her own company - let alone across companies. In addition, this physician typically does not have reliable data about how the inclusion or exclusion criteria, the clinical parameters that determine whether a given individual may participate in a clinical trial, will affect the number of patients eligible for the study.

A pharmaceutical research staff member typically translates portions of the trial protocol into a Case Report Form (CRF) manually using word-processor technology and personal experience with a limited number of previous trials. The combined cutting and pasting in both protocol and CRF development often results in redundant items or even irrelevant items being carried over from trial to trial. Data managers typically design and build database structures manually to capture the expected results. When the protocol is amended due to changes in FDA regulations, low accrual rates, or changing practices, as often occurs several times over the multiple years of a big trial, all of these steps are typically repeated manually.

At the trial site, which is often a physician's office, each step of the process from screening patients to matching the protocol criteria, through administering the required diagnostics and therapeutics, to collecting the data both internally and from outside labs, is usually done manually by individuals with another primary job (doctors and nurses seeing 'routine patients') and using paper based systems. The result is that patients who are eligible for a trial often are not recruited or enrolled, errors in following the trial protocol occur, and patient data are often either not captured at all, or are incorrectly transcribed to the CRF from hand written medical records, and are illegible. An extremely large percentage of the cost of a trial is consumed with data audit tasks such as resolving missing data, reconciling inconsistent data, data entry and validation. All of these tasks must be completed before the database can be "locked," statistical analysis can be performed and submission reports can be created.

## 2. Implementation

Once the trial is underway, data begins flowing back from multiple sites typically on paper forms. These forms routinely contain errors in copying data from source documents to CRFs.

5 Even without transcription errors, the current model of retrospective data collection is severely flawed. It requires busy investigators conducting multiple trials to correctly remember and apply the detailed rules of every protocol. By the time a clinical coordinator fills out the case report form the patient is usually gone, meaning that any data that was not collected or treatment protocol complexities that were not followed are generally unrecoverable. This occurs whether  
10 the case report form is paper-based or electronic. The only solution to this problem is point-of-care data capture, which historically has been impractical due to technology limitations.

Once the protocol is in place it often has to be amended. Reasons for changing the protocol include new FDA guidelines, amended dosing rules, and eligibility criteria that are found to be so restrictive that it is not possible to enroll enough patients in the trial. These "accrual  
15 delays" are among the most costly and time-consuming problems in clinical trials.

The protocol amendment process is extremely labor intensive. Further, since protocol amendments are implemented at different sites at different times, sponsors often don't know which protocol is running where. This leads to additional 'noise' in the resulting data and downstream audit problems. In the worst case, patients responding to an experimental drug may  
20 not be counted as responders due to protocol violations, but even count against the response rate under an intent-to-treat analysis. It is even conceivable that this purely statistical requirement could cause an otherwise useful drug to fail its trials.

Sponsors, or Contract Research Organizations (CROs) working on behalf of sponsors, send out armies of auditors to check the paper CRFs against the paper source documents. Many  
25 of the errors they find are simple transcription errors in manually copying data from one paper to the other. Other errors, such as missing data or protocol violations, are more serious and often unrecoverable.

### 3. Monitoring

The monitoring and audit functions are one of the most dysfunctional parts of the trial process. They consume huge amounts of labor costs, disrupt operations at trial sites, contribute to high turnover, and often involve locking the door after the horse has bolted.

5

### 4. Reporting

As information flows back from sites, the mountain of paper grows. The typical New Drug Application (NDA) literally fills a semi-truck with paper. The major advance in the past few years has the addition of electronic filing, but this is basically a series of electronic page copies of the same paper documents - it does not necessarily provide quantitative data tables or other tools to automate analysis.

10

### B. The Costs of Inefficiency

It can be seen that this complex manual process is highly inefficient and slow. And since each trial is largely a custom enterprise, the same thing happens all over again with the next trial. Turnover in the trials industry is also high, so valuable experience from trial to trial and drug to drug is often lost.

15

The net result of this complex, manual process is that despite accumulated experience, it is costing more to conduct each successive trial.

20

In addition to being slow and expensive, the current clinical trial process often hurts the market value of the resulting drug in two important ways. First, the FDA reviews drugs on an "intent to treat" basis. That means that every patient enrolled in a trial is included in the denominator (positive responders/total treated) when calculating a drug's efficacy. However, only patients who respond to treatment and comply with the protocol are included in the numerator as positive responders. Not infrequently, a patient responds to a drug favorably, but is actually counted as a failure due to significant protocol non-compliance. In rare cases, an entire

25

trial site is disqualified due to non-compliance. Non-compliance is often a result of preventable errors in patient management.

The second major way that the current clinical trial process hurts drug market value is that much of the fine grain detail about the drug and how it is used is not captured and passed from clinical development to marketing within a pharmaceutical company. As a result, virtually every pharmaceutical company has a second medical department that is a part of the marketing group. This group often repeats studies similar to those used for regulatory approval in order to capture the information necessary to market the drug effectively.

#### C. The Situation at Trial Sites

Despite the existence of a large number of clinical trials that are actively recruiting patients, only a tiny percentage of eligible patients are enrolled any clinical trial. Physicians, too, seem reluctant to engage in clinical trials. One study by the American Society of Clinical Oncology found that barriers to increased enrollment included restrictive eligibility criteria, large amount of required paperwork, insufficient support staff, and lack of sufficient time for clinical research.

Clinical trials consist of a complex sequence of steps. On average, a clinical trial requires more than 10 sites, enrolls more than 10 patients per site and contains more than 50 pages for each patient's case report form (data entry sheet). Given this complexity, delays are a frequent occurrence. A delay in any one step, especially in early steps such as patient accrual, propagates and magnifies that delay downstream in the sequence.

A significant barrier to accurate accrual planning is the difficulty trial site investigators have in predicting their rate of enrollment until after a trial as begun. Even experienced investigators tend to overestimate the total number of enrolled patients they could obtain by the end of the study. Novice investigators tend to overestimate recruitment potential by a larger margin than do experienced investigators, and with the rapid increase in the number of

investigators participating in clinical trials, the vast majority of current investigators have not had significant experience in clinical trials.

**D. Absence of Information Infrastructure**

5 Given the above state of affairs, one might expect that the clinical trials industry would be ripe for automation. But despite the desperate need for automation, remarkably little has been done.

10 While the pharmaceutical industry spends hundreds of millions of dollars annually on clinical information systems, most of this investment is in internal custom databases and systems within the pharmaceutical company; very little of this technology investment is at the physician office level. Each trial, even when conducted by the same company or when testing the same drug, is usually a custom collection of sites, procedures and protocols. More than half of trials are conducted for the pharmaceutical industry by Contract Research Organizations (CROs) using the same manual systems and custom physician networks.

15 The clinical trials information technology environment contributes to this situation. Clinical trials are information-intensive processes - in fact, information is their only product. Despite this, there is no comprehensive information management solution available. Instead there are many vendors, each providing tools that address different pieces of the problem. Many of these are good products that have a role to play, but they do not provide a way of integrating or managing information across the trial process.

20 The presently available automation tools include those that fall into the following major categories:

- Clinical data capture (CDC)
- 25 • Site-oriented trial management
- Electronic Medical Records (EMRs) with Trial-Support Features
- Trial Protocol design tools



- Site-sponsor matching services
- Clinical data management

5

Clinical Research Organizations (CROs) and Site Management Organizations (SMOs) also provide some information services to trial sites and sponsors.

### 1. Clinical Data Capture (CDC) Products

10

These products are targeted at trial sites, aiming to improve speed and accuracy of data entry. Most are rapidly moving to Web-based architectures. Some offer off-line data entry, meaning that data can be captured while the computer is disconnected from the Internet. Most companies can point to half a dozen pilot sites and almost no paying customers.

15

These products do not create an overall, start-to-finish, clinical trials management framework. These products also see "trial design" merely as "CRF design," ignoring a host of services and value that can be provided by a comprehensive clinical trials system. They also fail to make any significant advance over conventional methods of treating each trial as a "one-off" activity. For example, the companies offering CDC products continue to custom-design each CRF for each trial, doing not much more than substituting HTML code for printed or word-processor forms.

20

### 2. Site-Oriented Trial Management

25

These products are targeted at trial sites and trial sponsors, aiming to improve trial execution through scheduling, financial management, accrual, visit tracking. These products do not provide electronic clinical data entry, nor do they assist in protocol design, trial planning for sponsors, patient accrual or task management.

### 3. Electronic Medical Records (EMR) with Trial-Support Features

These products aim to support patient management of all patients, not just study patients, replacing most or all of a paper charting system. Some EMR vendors are focusing on particular disease areas, with KnowMed being a notable example in oncology.

These products for the most part do not focus specifically on the features needed to support clinical trials. They also require major behavior changes affecting every provider in a clinical setting, as well as requiring substantial capital investments in hardware and software. Perhaps because of these large hurdles, EMR adoption has been very slow.

### 4. Trial Protocol Design Tools

These products are targeted at trial sponsors, aiming to improve the protocol design and program design processes using modeling and simulation technologies. One vendor in this segment, PharSight, is known for its use of PK/PD (pharmacokinetic/pharmacodynamic) modeling tools and is extending its products and services to support trial protocol design more broadly.

None of the companies offering trial protocol design tools provide the host of services and value that can be provided by a comprehensive clinical trials system.

### 5. Trial Matching Services

Some recent Web-based services aim to match sponsors and sites, based on a database of trials by sponsor and of sites' patient demographics. A related approach is to identify trials that a specific patient may be eligible for, based on matching patient characteristics against a database of eligibility criteria for active trials. This latter functionality is often embedded in a disease-specific healthcare portal such as cancerfacts.com.

### 6. Clinical Data Management

Two well-established products, Domain ClinTrial and Oracle Clinical, support the back-end database functionality needed by sponsors to store the trial data coming in from CRFs.

These products provide a visit-specific way of storing and querying study data. The protocol sponsor can design a template for the storage of such data in accordance with the protocol's visit schema, but these templates are custom-designed for each protocol. These products do not provide protocol authoring or patient management assistance.

5

#### **7. Statistical Analysis**

The SAS Institute (SAS) has defined the standard format for statistical analysis and FDA reporting. This is merely a data format, and does not otherwise assist in the design or execution of clinical trial protocols.

#### **8. Site Management Organizations (SMOs)**

SMOs maintain a network of clinical trial sites and provide a common Institutional Review Board (IRB) and centralized contracting/invoicing. SMOs have not been making significant technology investments, and in any event, do not offer trial design services to sponsors.

#### **9. Clinical Research Organizations (CROs)**

CROs provide, among other services, trial protocol design and execution services. But they do so on substantially the same model as do sponsors: labor-intensive, paper-based, slow, and expensive. CROs have made only limited investments in information technology.

#### **E. The Need for a Comprehensive Clinical Trials System**

It can be seen that the current information model for clinical trials is highly fragmented. This has led to high costs, "noisy" data, and long trial times. Without a comprehensive, service-oriented information solution it is very hard to get away from the current paradigm of paper, faxes and labor-intensive processes. And it has become clear that simply "throwing more bodies"

at trials will not produce the required results, particularly as trial throughput demands increase. A new, comprehensive model is required.

### SUMMARY OF THE INVENTION

5 According to the invention, roughly described, clinical trials are defined, managed and evaluated according to an overall end-to-end system solution which starts with the creation of protocol meta-models by a central authority, and ends with the conduct of trials by clinical sites, who then report back electronically for near-real-time monitoring by study sponsors and for analysis by the central authority. The central authority first creates protocol meta-models, one  
10 for each of several different disease categories, and makes them available to protocol designers. Each meta-model includes a short list of preliminary patient eligibility attributes which are appropriate for a particular disease category. The protocol designer chooses the meta-model and preliminary eligibility list appropriate for the relevant disease category, and encodes the clinical trial protocol, including eligibility and patient workflow, within the selected meta-model. The  
15 resulting protocol database is stored together with databases of other protocols in the same and different disease categories, in a library of protocol databases maintained by the central authority. Sponsors and individual clinical sites have access to only the particular protocols for which they are authorized.

20 Study sites optionally use a two-stage screening procedure in order to identify clinical studies for which individual patients are eligible, and patients who are eligible for individual clinical studies. The study sites make reference to the protocol databases to which they have access in the protocol database library in order to make these determinations. In one embodiment the data that is gleaned from patients being screened is retained in a patient-specific database of patient attributes, or in other embodiments the data can be stored anonymously or discarded after  
25 screening. Once a patient is enrolled into a study, the protocol database indicates to the clinician exactly what tasks are to be performed at each patient visit. These tasks can include both patient management tasks, such as administering a drug or taking a measurement, and also data

management tasks, such as completing and submitting a particular CRF. The workflow graph embedded in the protocol database advantageously also instructs the proper time for the clinician to obtain informed consent from a patient during the eligibility screening process, and when to perform future tasks, such as the acceptable date range for the next patient visit.

5 The system keeps track of the progress of the patient and the clinician through the workflow graph of a particular protocol. The system reports this information to study sponsors, who can then monitor the progress of an overall clinical trial in near-real-time, and to the central authority which can then generate performance metrics. Advantageously, a common controlled medical terminology database is used by all components of the system in order to ensure  
10 consistent usage of medical terminology by all the participants. The protocol database is advantageously used also to drive other kinds of problem-solving methods, such as an accrual simulation tool and a Find-Me-Patients tool.

### BRIEF DESCRIPTION OF THE DRAWINGS

15 The invention will be described with respect to specific embodiments thereof, and reference will be made to the drawings, in which:

Fig. 1 is a symbolic block diagram illustrating significant aspects of a clinical trials management system and method incorporating features of the invention.

20 Figs. 2-8 are screen shots of an example for an Intelligent Clinical Protocol (iCP) database.

Fig. 9 is a flow chart detail of the step of creating iCPs in Fig. 1.

Fig. 10 is a flow chart of an optional method for a protocol author to establish patient eligibility criteria.

25 Figs. 11-25 are screen shots of screens produced by Protégé 2000, and will help illustrate the relationship between a protocol meta-model and an example individual clinical trial protocol.

Fig. 26 is a flow chart detail of step 122 (Fig. 1).

## DETAILED DESCRIPTION

Fig. 1 is a symbolic block diagram illustrating significant aspects of a clinical trials management system and method incorporating features of the invention. In the figure, solid arrows indicate process flow, whereas broken arrows indicate information flow. In broad summary, the system is an end-to-end solution which starts with the creation of protocol meta-models by a central authority, and ends with the conduct of trials by clinical sites, who then report back electronically for near-real-time monitoring by study sponsors, for analysis by the central authority, and for use by study sponsors in identifying promising sites for future studies. As used herein, a "clinical site" can be physically at a single or multiple locations, but conducts clinical trials as a single entity. The term also includes SMOs.

Referring to Fig. 1, the central authority initially creates one or more protocol meta-models (step 110) for use in facilitating the design of clinical trial protocols. Each meta-model can be thought of as a set of building blocks from which particular protocols can be built. Preferably, the central authority creates a different meta-model for each of several disease classifications, with the building block in each meta-model being appropriate to that disease classification. In an embodiment, a meta-model is described in terms of object oriented design. The building blocks are represented as object classes, and an individual protocol database contains instances of the available classes.

The building blocks contained in a meta-model include the different kinds of steps that might be required in a trial protocol workflow, such as, for example, a branching step, an action step, a synchronization step, and so on. The available action steps for a meta-model directed to breast cancer trials might differ from the available action steps in a meta-model directed to prostate cancer trials, for example, by making available only those kinds of steps which might be appropriate for the particular disease category. For example, a step of brachytherapy might be available in the prostate cancer meta-model, but not in the breast cancer meta-model; and a step of mammography might be available in the breast cancer meta-model, but not in the prostate cancer meta-model.

The meta-models also include lists, again appropriate to the particular disease category, within which a protocol designer can define preliminary criteria for the eligibility of patients for a particular study. These preliminary eligibility criteria lists do not preclude a protocol designer from building further eligibility criteria into any particular clinical trial protocol. As set forth in more detail below, the options available in the lists of preliminary eligibility criteria are intentionally limited in number so as to facilitate the building of a large database of potential patients for studies within the particular disease area. At the same time, however, it is also desirable that the options be numerous or narrow enough in order to provide a good first cut of eligible patients. In order to best satisfy these two competing goals, it is desirable that an expert or a team of experts knowledgeable about the particular disease category of a particular meta-model be heavily involved in the development of the preliminary eligibility criteria lists for the particular meta-model. In addition, because of the difficulty and length of time required to develop a large database of potential patients, it is further desirable that once the eligibility criteria options are established for a particular meta-model, they do not change except as absolutely necessary. Such changes might be mandated as a result of improved understanding of a disease, for example, and are rigorously managed throughout the overall system of Fig. 1.

Table I sets forth example Preliminary Eligibility Criteria lists for five disease categories, specifically breast cancer, small cell lung cancer, non-small cell lung cancer, colorectal cancer and prostate cancer. As can be seen, each list includes a small number of patient attributes, each with a set of available choices from which the protocol designer can choose, in order to encode preliminary eligibility criteria for a particular protocol. The protocol meta-model for breast cancer, for example, includes the list of attributes and the list of available choices for each attribute, as shown in the row of the table for "Breast Cancer."

**TABLE I**  
**Example Preliminary Eligibility Criteria Lists**

<b>Disease</b>	<b>QS attribute</b>	<b>Choices</b>
Breast cancer	Current Stage	O, I, II (IIA, IIB), III (IIIA, IIIB), IV
	Prior Chemo	None, Neoadj/Adj, Tx Adv Disease
	Prior RT	None, Primary tumor, Metastatic Dz
	Prior Surgery	Y,N
	Prior Hormonal	None, Neoadj/Adj, Tx Adv Disease
Lung cancer, small cell	Current Stage	Limited, Extensive
	Prior Chemo	None, Neoadj/Adj, Tx Adv Disease
	Prior RT	None, Primary tumor, Metastatic Dz
	Prior Surgery	Y,N
Lung cancer, non-small cell	Current Stage	O, I (IA, IB), II (IIA, IIB), IIIA, IIIB, IV
	Prior Chemo	None, Neoadj/Adj, Tx Adv Disease
	Prior RT	None, Primary tumor, Metastatic Dz
	Prior Surgery	Y,N
Colorectal cancer	Current Stage	O, I, II, III, IV
	Prior Chemo	None, Neoadj/Adj, Tx Adv Disease
	Prior RT	None, Primary tumor, Metastatic Dz
	Prior Surgery	Y,N
Prostate cancer	Metastases	Y,N
	Primary Tumor	N/A, T0, T1a, T1b, T1c, T2 (T2a, T2b), T3 (T3a, T3b), T4
	Nodes	N/A, N0, N1
	Prior Chemo	None, Neoadj/Adj, Tx Adv Disease
	Prior RT	None, Primary tumor, Metastatic Dz
	Prior Surgery	Y,N
	Prior Hormonal	None, Neoadj/Adj, Tx Adv Disease

In the embodiment illustrated by Table I, the designer encodes preliminary eligibility criteria by assigning one of the available choices to each of at least a subset of the attributes in



the selected list. Each "criterion" is defined by an attribute and its assigned value, so that a patient satisfies the criterion only if the patient has the specified value for that attribute. Each criterion is then classified either as an "inclusion" criterion or an "exclusion" criterion; a patient must satisfy all the inclusion criteria and none of the exclusion criteria in order to pass preliminary eligibility.

The logic of the preliminary eligibility criteria is capable of many variations in different embodiments. In one embodiment, for example, the designer is permitted to assign more than one of the available values to a given attribute. In this case, a patient who has any of the assigned values for the given attribute satisfies the criterion. In another embodiment, one or more of the available values for a given attribute can be specified as a numeric range, and the designer can assign any value or sub-range of values within that range to the attribute. In addition, the condition for satisfying the criterion can be specified by the designer to be something other than equality (e.g., "attribute having a value less than N", or "attribute having a value greater than N".) Speaking more generally, each criterion is defined by an attribute and a "condition", and the patient must satisfy the condition with respect to that attribute in order to satisfy the criterion.

Applicants believe that one of the reasons why in the past it has been difficult to create tools which are useful throughout the clinical trials design, execution and analysis processes, as shown in Fig. 1, is that each tool presently available maintains its data in a different format, and because there is no consistency among the different tools regarding such important questions as what is meant by a particular medical term according to which the data is maintained. The problem of consistent medical terminology is one that extends well beyond the field of clinical trials, and has spawned significant academic discourse within the field of medical informatics. As examples, see the entire November 1998 issue of "Methods of Information in Medicine", Vol. 37, No. 4-5, which is incorporated herein by reference in its entirety. See in particular the following articles: "Editorial: the Concepts of Language and the Language of Concepts", by J.J. Cimino (p. 311); "Desiderata for Controlled Medical Vocabularies in the Twenty-First Century", by J.J. Cimino (pp. 394-403); "Concept-Oriented Standardization and Statistics-Oriented

Classification: Continuing the Classification Versus Nomenclature Controversy", by J. Ingenerf and W. Giere (pp. 527-539); and "Standards to Support Development of Terminological Systems for Health Care Telematics", by A. Rossi Mori, F. Consorti and E. Galeazzi (pp. 551-563). See also, Broverman, "Overview of Controlled Clinical Terminologies", slides for CME course, August 18, 1999, incorporated herein by reference.

The work on terminology consistency problems has yielded a number of different databases of medical concepts, terms and attributes, none of which have been intended for use in the field of clinical trial protocols but most of which can provide some benefit in that field. These terminology databases, sometimes referred to herein as "controlled medical terminologies" (CMTs), include interface terminologies intended to support navigation for structured data entry, reference terminologies intended to support aggregation and analysis of medical data, and administrative terminology intended to support reporting and billing requirements. Interface terminologies, which include such terminologies as MEDCIN, Oceania CKB, and Purkinjie, are patient-dependent terminologies, are easily navigable to support workflow, and provide groupings of observations by context. Reference terminologies, which include SNOMED, LOINC, Meddra, and Read, are patient-independent terminologies. They are intended to completely cover a domain of discourse. They include definitions of logical concepts and are organized into hierarchies of semantic classifications. Administrative terminologies, which include ICD-9 and CPT-4, are also patient-independent. They are designed primarily for reporting and billing requirements and do not contain formal definitions. They are organized hierarchically, but only to a limited extent. In addition to the classification of CMTs as interface terminologies, reference terminologies or administrative terminologies, some of the existing CMTs (such as SNOMED) are better at covering medical diagnoses than others, and others (such as CPT-4) are better at covering the domain of medical procedures. Meddra is better at covering the field of adverse events reporting, having been developed by the Food and Drug Administration primarily for post-market surveillance of drug usage experience. All of the above-mentioned CMTs are incorporated by reference herein.

10 The overall clinical trials process illustrated in Fig. 1 is performed by a wide variety of  
different people, all of whom might have different understandings about the meaning of various  
concepts, terms and attributes. Therefore, in order for all the different steps and tools to work  
well together, the system of Fig. 1 takes advantage of a CMT 112 wherever possible. For  
5 example, most if not all of the concepts, terms and attributes which are used in the workflow task  
building blocks and patient eligibility criteria options made available in the meta-models produced  
in step 110, are entries in the CMT 112. In one embodiment, a meta-model contains copies of  
the relevant entries from the CMT 112, whereas in another embodiment, the meta-models  
contain merely pointers to the relevant entries in CMT 112. A meta-model designer or protocol  
author preferably uses a CMT browser to select desired CMT entries for inclusion in a meta-  
model or protocol.

15 The CMT 112 can be any of the CMTs presently existing, or can be a new CMT  
altogether. Preferably, CMT 112 includes entries developed from several of the different  
presently existing CMTs, as well as additional entries which are appropriate for clinical trial  
protocols, and for the particular disease categories covered by the various meta-models produced  
in step 110. Preferably CMT 112 is organized hierarchically by concept. Each concept includes  
pointers to one or more terms which describe the concept synonymously. For example, the  
concept "hypertension" might have pointers to the terms "hypertension", "elevated blood  
pressure", and "high blood pressure". Each concept also includes pointers to one or more  
20 attributes. The "hypertension" concept, for example, might include a pointer to an attribute,  
"diastolic blood pressure greater than 95mm Hg," where "blood pressure" is itself a concept and  
is represented by a pointer back to the list of concepts.

25 The step 110 of creating protocol meta-models is performed using a meta-model  
authoring tool. Protégé 2000 is an example of a tool that can be used as a meta-model authoring  
tool. Protégé 2000 is described in a number of publications including William E. Grosso, et. al.,  
"Knowledge Modeling at the Millennium (The Design and Evolution of Protégé-2000)," SMI  
Report Number: SMI-1999-0801 (1999), available at <http://smi-web.stanford.edu/>

pubs/SMI\_Abstracts/SMI-1999-0801.html, visited 01/01/2000, incorporated by reference herein. In brief summary, Protégé 2000 is a tool that helps users build other tools that are custom-tailored to assist with knowledge-acquisition for expert systems in specific application areas. It allows a user to define "generic ontologies" for different categories of endeavor, and then to  
5 define "domain-specific ontologies" for the application of the generic ontology to more specific situations. In many ways, Protégé 2000 assumes that the different generic ontologies differ from each other by major categories of medical endeavors (such as medical diagnosis versus clinical trials), and the domain-specific ontologies differ from each other by disease category. In the present embodiment, however, all ontologies are within the category of medical endeavor known as clinical trials and protocols. The different generic ontologies correspond to the different meta-models produced in step 110 (Fig. 1), which differ from each other by disease category. In this sense, the generic ontologies produced by Protégé in step 110 are directed to a much more  
10 specific domain than those produced in other applications of Protégé 2000.

Since the meta-models produced in step 110 include numerous building blocks as well as many options for patient eligibility criteria, a wide variety of different kinds of clinical trial  
15 protocols, both simple and complex, can be designed. These meta-models are provided to clinical trial protocol designers who use them, preferably again with the assistance of Protégé 2000, to design individual clinical trial protocols in step 114.

In step 114 of Fig. 1, a protocol designer desiring to design a protocol for a clinical trial  
20 in a particular disease category, first selects the appropriate meta-model and then uses the authoring tool to design and memorialize the protocol. As in step 110, one embodiment of the authoring tool for step 114 is based on Protégé 2000. The output of step 114 is a database which contains all the significant required elements of a protocol. This database is sometimes referred to herein as an Intelligent Clinical Protocol (iCP) database, and provides the underlying logical  
25 structure for driving much of the processes that take place in the remainder of Fig. 1.

Conceptually, an iCP database is a computerized data structure that encodes most significant aspects of a clinical protocol, including eligibility criteria, randomization options,

treatment sequences, data requirements, and protocol modifications based on patient outcomes or complications. The iCP structure can be readily extended to encompass new concepts, new drugs, and new testing procedures as required by new drugs and protocols. The iCP database is used by most software modules in the overall system to ensure that all protocol parameters, treatment decisions, and testing procedures are followed.

The iCP database can be thought of as being similar to the CAD/CAM tools used in manufacturing. For example, a CAD/CAM model of an airplane contains objects which represent various components of an airplane, such as engines, wings, and fuselage. Each component has a number of additional attributes specific to that component - engines have thrust and fuel consumption; wings have lift and weight. By constructing a comprehensive model of an airplane, numerous different types of simulations can be executed using the same model to ensure consistent results, such as flight characteristics, passenger/revenue projections, maintenance schedules. And finally, the completed CAD/CAM simulations automatically produced drawings and manufacturing specifications to accelerate actual production. While an iCP database differs from the CAD/CAM model in important ways, it too provides a comprehensive model of a clinical protocol so as to support consistent tools created for problems such as accrual, patient screening and workflow management. By using a comprehensive model and a unifying standard vocabulary, all tools behave according to the protocol specifications.

As used herein, the term "database" does not necessarily imply any unity of structure. For example, two or more separate databases, when considered together, still constitute a "database" as that term is used herein.

As described in more detail below, the iCP data structures can be used by multiple tools to ensure that the tool performs in strict compliance with the clinical protocol requirements. For example, a patient recruitment simulation tool can use the eligibility criteria encoded into an iCP data structure, and a workflow management tool uses the visit-specific task guidelines and data capture requirements encoded into the iCP data structure. The behavior of all such tools will be consistent with the protocol because they all use the same iCP database.

Many clinical systems provide a "dumb database" for patient data, but offer no intelligence, no automation. While these systems may offer some efficiency benefits compared to paper systems, they are incapable of driving workflow management, sophisticated data validation or recognizing protocol-critical patterns in patient data (e.g. a toxic response to a drug that should trigger a modification to the treatment). A few systems have used rule-based expert systems or other technologies to deliver more intelligence to clinicians, but these have encountered significant problems: huge up-front modeling costs and ongoing maintenance costs; unpredictable system behavior over time; and an inability to reuse knowledge content or software components. So the choices available for clinical investigators have been poor: use paper, use an electronic file cabinet with no intelligence, or build a custom intelligent system for each trial. The use of an iCP database and a variety of tools designed to be driven by an iCP database overcomes many of the deficiencies with the prior art options.

The iCP database is used to drive all downstream "problem solvers" such as electronic case report forms, and assures that those applications are revised automatically as the protocol changes. This assures protocol compliance. The iCP authoring tool draws on external knowledge bases to help trial designers, and creates a library of re-usable protocol "modules" that can be incorporated in new trials, saving time and cost and enabling a clinical trial protocol design process that is more akin to customization than to the current "every trial unique" model.

Figs. 11-25 are screen shots of screens produced by Protégé 2000, and will help illustrate the relationship between a protocol meta-model and an example individual clinical trial protocol. Fig. 11 is a screen shot illustrating the overall class structure in the left-hand pane 1110. Of particular interest to the present discussion is the class 1112, called "FastTrackClass" and the classes under class 1112. FastTrackClass 1112 and those below it represent an example of a protocol meta-model. This particular meta-model is not specific to a single disease category, but one that is specific to a single disease category is described below with respect to Figs. 2-8.

The right-hand pane 1114 of the screen shot of Fig. 11 sets forth the various slots that have been established for a selected one of the classes in the left-hand pane 1110. In the image

of Fig. 11, the "protocol" class 1116, a subclass of FastTrackClass 1112, has been selected (as indicated by the shading). In the right-hand pane 1114, specifically in the window 1118, the individual slots for protocol class 1116 are shown. Only those indicated by a shaded "S" are pertinent to the present discussion; those indicated by an unshaded "S" are more general and not important for an understanding of the invention. It can be seen that several of the slots in the window 1118 contain "facets" which, for some slots, define a limited set of "values" that can be stored in the particular slot. For example, the slot "quickScreenCriterion" can take on only the specific values "prostate cancer," "colorectal cancer," "breast cancer," etc. These are the only disease categories for which quickScreenCriteria had been established at the time the screen shot of Fig. 11 was taken.

Fig. 12 is a screen shot of a particular instance of class "protocol" in Fig. 11, specifically a protocol object having identifier CALGB 9840. It can be seen that each of the slots defined for protocol class 1116 has been filled in with specific values in the protocol class object instance of Fig. 12. Whereas Fig. 11 illustrates an aspect of a clinical trial protocol meta-model, Fig. 12 illustrates the top-level object of an actual iCP designated CALGB 9840. Of particular note, it can be seen that for the iCP CALGB 9840, the slot "quickScreenCriterion" 1120 (Fig. 11) has been filled in by the protocol author as "Breast Cancer" (item 1210 in Fig. 12), which is one of the available values 1122 for the quickScreenCriterion slot 1120 in Fig. 11. In addition, the protocol author has also filled in "CALGB 9840 Eligibility Criteria", an instance of EligibilityCriteriaSet class 1124, for an EligibilityCriteriaSet slot (not shown in Fig. 11) of the protocol class object. Essentially, therefore, the protocol class object of Fig. 12 includes a pointer to another object identifying the "further eligibility criteria" for iCP CALGB 9840.

Fig. 13 illustrates in the right-hand pane 1310 the slots defined in the protocol meta-model for the class "EligibilityCriteriaSet" 1124. Of particular note is that an EligibilityCriteriaSet object will include both exclusion criteria (slot 1312) and inclusion criteria (slot 1314). It can be seen from Fig. 13 that the values that can be placed in slots 1312 and 1314 are objects of the class "EligibilityCriterion" 1126. It will be appreciated that in a different

embodiment, other structural organizations for maintaining the same information are possible, such as a single list including all patient eligibility criteria, and flags indicating whether each criterion is an inclusion criterion or an exclusion criterion.

Fig. 14 illustrates in the right-hand pane 1410 the slots which can be filled in for objects of the class "EligibilityCriterion". As can be seen, these slots are merely for descriptive text strings, primarily a slot 1412 for a long description and a slot 1414 for a short description.

Fig. 15 illustrates the instance of the EligibilityCriteriaSet class which appears in the CALGB 9840 iCP. It can be seen that the object contains a list of inclusion criteria and a list of exclusion criteria, each criterion of which is an instance of the EligibilityCriterion class 1126. One of such instances 1510 is illustrated in Fig. 16. Only the short description 1610 and the long description 1612 have been entered by the protocol author.

An iCP, in addition to containing a pointer (1210 in Fig. 12) to the relevant set of quickScreenCriteria, and also identifying (1212) further eligibility criteria, also contains the protocol workflow in the form of patient visits, management tasks to take place during a visit, and transitions from one visit to another. The right-hand pane 1710 of Fig. 17 illustrates the slots available for an object instance of the class "visit" 1128. It can be seen that in addition to a slot 1712 for possible visit transitions, the Visit class also includes a slot 1714 for patient management tasks as well as another slot 1716 for data management tasks. In other words, a clinical trial protocol prepared using this clinical trial protocol meta-model can include instructions to clinical personnel not only for patient management tasks (such as administer certain medication or take certain tests), but also data management tasks (such as to complete certain CRFs).

Fig. 18 illustrates a particular instance of visit class 1128, which is included in the CALGB 9840 iCP. As can be seen, it includes a window 1810 containing the possible visit transitions, a window 1812 containing the patient management tasks, and a window 1816 showing the data management tasks for a particular visit referred to as "Arm A treatment visit".

The data management tasks and patient management tasks are all instance of the



"ManagementTask" class 1130 (Fig. 11), the slots of which are set forth in the right-hand pane 1910 of Fig. 19. As with the EligibilityCriterion class 1126 (Fig. 14), the slots available to a protocol author in a ManagementTask object are mostly text fields.

Fig. 20 illustrates the ManagementTask object 1816 (Fig. 18), "Give Arm A Paclataxel Treatment." Similarly, Fig. 21 illustrates the ManagementTask object 1818, "Submit Form C-116". The kinds of data management tasks which can be included in an iCP according to the clinical trial protocol meta-model include, for example, tasks calling for clinical personnel to submit a particular form, and a task calling for clinical personnel to obtain informed consent.

Returning to Fig. 17, the values that a protocol author places in the slot 1712 of a visit class 1128 object are themselves instances of VisitToVisitTransition class 2210 (Fig. 22) in the meta-model. The right-hand pane 2212 shows the slots which are available in an object of the VisitToVisitTransition class 2210. As can be seen, it includes a slot 2214 which points to the first visit object of the transition, another slot 2216 which points to a second visit object of the transition, and three slots 2218, 2220 and 2222 in which the protocol author provides the minimum, maximum and preferred relative time of the transition. Fig. 23 shows the contents of a VisitToVisitTransition object 1818 (Fig. 18) in the CALGB 9840 iCP.

In addition to being kept in the form of Visit objects, management task objects and VisitToVisitTransition objects, the protocol meta-model also allows an iCP to keep the protocol schema in a graphical or diagrammatic form as well. In fact, it is the graphical form that protocol authors typically use, with intuitive drag-and-drop and drill-down behaviors, to encode clinical trial protocols using Protégé 2000. In the protocol meta-model, a slot 1134 is provided in the Protocol object class 1116 for pointing to an object of the ProtocolSchemaDiagram class 1132 (Fig. 11). Fig. 24 shows the slots available for ProtocolSchemaDiagram class 1132. As can be seen, they include a slot 2410 for diagrammatic connectors, and another slot 2412 for diagram nodes. The diagram connectors are merely the VisitToVisitTransition objects described previously, and the diagram nodes are merely the Visit objects described previously. Fig. 25 illustrates the ProtocolSchemaDiagram object 1214 (Fig. 12) in the CALGB 9840 iCP. It can

be seen that the entire clinical trial protocol schema is illustrated graphically in pane 2510, and the available components of the graph (connector objects 2512 and visit objects 2514) are available in pane 1516 for dragging to desired locations on the graph.

Figs. 2-8 are screen shots of another example iCP database, created and displayed by Protégé 2000 as an authoring tool. This iCP encodes clinical trial protocol labeled CALGB 49802, and differs from the CALGB 9840 iCP in that CALGB 49802 was encoded using a starting meta-model that was already specific to a specific disease area, namely cancer. It will be appreciated that in other embodiments, the meta-models can be even more disease specific, for example meta-models directed specifically to breast cancer, prostate cancer and so on.

Fig. 2 is a screen shot of the top level of the CALGB 49802 iCP database. The screen shot sets forth all of the text fields of the protocol, as well as a list 210 of patient inclusion criteria and a list 212 of patient exclusion criteria.

Fig. 3 is a screen shot of the Management\_Diagram class object for the iCP, illustrating the workflow diagram for the clinical trial protocol of Fig. 2. The workflow diagram sets forth the clinical algorithm, that is, the sequence of steps, decisions and actions that the protocol specification requires to take place during the course of treating a patient under the particular protocol. The algorithm is maintained as sets of tasks organized as a graph 310, illustrated in the left-hand pane of the screen shot of Fig. 3. The protocol author adds steps and/or decision objects to the graph by selecting the desired type of object from the palette 312 in the right-hand pane of the screen shot of Fig. 3, and instantiating them at the desired position in the graph 310. Buried beneath each object in the graph 310 are fields which the protocol designer completes in order to provide the required details about each step, decision or action. The user interface of the authoring tool allows the designer to drill down below each object in the graph 310 by double-clicking on the desired object. The Management\_Diagram object for the iCP also specifies a First Step (field 344), pointing to Consent & Enroll step 314, and a Last Step (field 346), which is blank.

Referring to the graph 310, it can be seen that the workflow diagram begins with a "Consent & Enroll" object 314. This step, which is described in more detail below, includes sub-steps of obtaining patient informed consent, evaluating the patient's medical information against the eligibility criteria for the subject clinical trial protocol, and if all such criteria are satisfied, enrolling the patient in the trial.

After consent and enrollment, step 316 is a randomization step. If the patient is assigned to Arm 1 of the protocol (step 318), then workflow continues with the "Begin CALGB 49802 Arm 1" step object 320. In this Arm, in step 322, procedures are performed according Arm 1 of the study, and workflow continues with the "Completed Therapy" step 324. If in step 318 the patient was assigned Arm 2, then workflow continues at the "Begin CALGB 49802 Arm 2" step 326. Workflow then continues with step 328, in which the procedures of protocol Arm 2 are performed and, when done, workflow continues at the "Completed Therapy" scenario step 324.

After step 324, workflow for all patients proceeds to condition\_step "ER+ or PR+" step 330. If a patient is neither estrogen-receptor positive nor progesterone receptor positive, then the patient proceeds to a "CALGB 49802 long-term follow-up" sub-guideline object step 332. If a patient is either estrogen-receptor positive or progesterone receptor positive, then the patient instead proceeds to a "Post-menopausal?" condition\_step object 334. If the patient is post-menopausal, then the patient proceeds to a "Begin Tamoxifen" step 336, and thereafter to the long-term follow-up sub-guideline 332.

If in step 334, the patient is not post-menopausal, then workflow proceeds to a "Consider Tamoxifen" choice\_step object 338. In this step, the physician using clinical judgment determines whether the patient should be given Tamoxifen. If so (choice object 340), then the patient continues to the "Begin Tamoxifen" step object 336. If not (choice object 342), then workflow proceeds directly to the long-term follow-up sub-guideline object 332. It will be appreciated that the graph 310 is only one example of a graph that can be created in different embodiments to describe the same overall protocol schema. It will also be appreciated that the library of object

classes 312 could be changed to a different library of object classes, while still being oriented to protocol-directed clinical studies.

Fig. 4 is a screen shot showing the result of "drilling down" on the "Consent & Enroll" step 314 (Fig. 3). As can be seen, Fig. 4 contains a sub-graph (which is also considered herein to be a "graph" in its own right) 410. The Consent & Enroll step 314 also contains certain text fields illustrated in Fig. 4 and not important for an understanding of the invention.

As can be seen, graph 410 begins with a "collect pre-study variables 1" step object 410, in which the clinician is instructed to obtain certain patient medical information that does not require informed consent. Step 412 is an "obtain informed consent" step, which includes a data management task instructing the clinician to present the study informed consent form to the patient and to request the patient's signature. In another embodiment, the step 412 might include a sub-graph which instructs the clinician to present the informed consent form, and if it is not signed and returned immediately, then to schedule follow-up reminder telephone calls at future dates until the patient returns a signed form or declines to participate.

After informed consent is obtained, the sub-graph 410 continues at step object 414, "collect pre-study variable 2". This step instructs the clinician to obtain certain additional patient medical information required for eligibility determination. If the patient is eligible for the study and wishes to participate, then the flow continues at step object 416, "collect stratification variables". The flow then continues at step 418, "obtain registration I.D. and Arm assignment" which effectively enrolls the patient in the trial.

Fig. 5 is a detail of the "Collect Stratification Variables" step 416 (Fig. 4). As can be seen, it contains a number of text fields, as well as four items of information that the clinician is to collect about the subject patient. When the clinical site protocol management software reaches this stage in the workflow, it will ask the clinician to obtain these items of information about the current patient and to record them for subsequent use in the protocol. The details of the "Collect pre-study variables" 1 and 2 steps 410 and 414 (Fig. 4) are analogous, except of course the specific tasks listed are different.

Fig. 6 is a detail of the "CALGB 49802 Arm 1" sub-guideline 332 (Fig. 3). As in Fig. 4, Fig. 6 includes a sub-graph (graph 610) and some additional information fields 612. The additional information fields 612 include, among other things, an indication 614 of the first step 618 in the graph, and an indication 616 of the last step 620 of the graph.

Referring to graph 610, the arm 1 sub-guideline begins with a "Decadron pre-treatment" step object 618. The process continues at a "Cycle 1; Day 1" object 622 followed by a choice\_object 624 for "Assess for treatment." The clinician may make one of several choices during step 624 including a step of delaying (choice object 626); a step of calling the study chairman (choice object 628); a step of aborting the current patient (choice object 630); or a step of administering the drug under study (choice object 632). If the clinician chooses to delay (object 626), then the patient continues with a "Reschedule next attempt" step 634, followed by another "Decadron pre-treatment" step 618 at a future visit. If in step 624 the clinician chooses to call the study chairman (object 628), then workflow proceeds to choose\_step object 636, in which the study chair makes an assessment. The study chair can choose either the delay object 626, the "Give Drug" object 632, or the Abort object 630.

If either the clinician (in object 624) or the study chair (in object 636) chooses to proceed with the "Give Drug" object 632, then workflow proceeds to choice\_step object 638 at which the clinician assesses the patient for dose attenuation. In this step, the clinician may choose to give 100% dose (choice object 640) or to give 75% dose (choice object 642). In either case, after dosing, the clinician then performs "Day 8 Cipro" step object 620. That is, on the 8<sup>th</sup> day, the patient begins a course of Ciprofloxacin (an antibiotic).

Without describing the objects in the graph 610 individually, it will be understood that many of these objects either are themselves specific tasks, or contain task lists which are associated with the particular step, visit or decision represented by the object.

Fig. 7 is a detail of the long term follow-up object 332 (Fig. 3). As mentioned in field 710, the first step in the sub-graph 712 of this object is a long term follow-up visit scenario visit object 714. That is, the sub-guideline illustrated in graph 712 is executed on each of the patient's

long-term follow-up visits. As indicated in field 724, the long term follow-up step 332 (Fig. 3) continues until the patient dies.

Object 716 is a case\_object which is dependent upon the patient's number of years post-treatment. If the patient is 1-3 years post-treatment, then the patient proceeds to step object 718, which among other things, schedules the next visit in 3-4 months. If the patient is 4-5 years post-treatment, then the patient proceeds to step object 720, which among other things, schedules the next patient visit in 6 months. If the patient is more than 5 years post-treatment, then the patient proceeds to step object 722, which among other things, schedules the next visit in one year. Accordingly, it can be seen that in the sub-guideline 712, different tasks are performed if the patient is less than 3 years out from therapy, 4-5 out from therapy, or more than 5 years out from therapy. Beneath each of the step objects 718, 720 and 722 are additional workflow tasks that the clinician is required to perform at the current visit.

Fig. 8 is an example detail of one of the objects 718, 720 or 722 (Fig. 7). It includes a graph 810 which begins with a CALGB 49802 f/u visit steps" consultation\_branch object 812, followed by seven elementary\_action objects 814 and 816a-f (collectively 816). Each of the consultation\_action objects 814 and 816 includes a number of workflow tasks not shown in the figures. It can be seen from the names of the objects, however, that the workflow tasks under object 814 are to be performed at every follow-up visit, whereas the workflow tasks under objects 816 are to be performed only annually.

As mentioned, the iCP facilitates integration of all the tools in the overall system of Fig. 1. Integration and especially consistency of language, is facilitated also by the use of CMT 112 (Fig. 1). As mentioned, a CMT provides for a consistent set of concepts, terms, and attributes which allow different tools to "understand" the same structures. CMT concepts include treatments, visits, and laboratory tests, for example. CMT terms include cytoxan, neutropenia, and WBC counts, for example. CMT attributes specify, for example, that cytoxan is a drug which has a dosage and can cause specific side effects, neutropenia is a clinical state described by a series of low white blood cell counts which occur together over a short period of time, and WBC

counts are blood test results which have a certain range and are a measurement whose validity persists for only a few days. Embodiments described herein use CMT to describe patient eligibility criteria (both preliminary eligibility criteria and further eligibility criteria), and also to describe the individual workflow tasks required of the clinician according to the protocol schema represented in the iCP. In one embodiment, the use of CMT is optional, whereas in another embodiment, the use of CMT is enforced. CMT is used in embodiments described herein also to describe patient medical status information that the clinician obtains and records. No existing clinical trials tool makes use of CMT as described herein.

Returning to Fig. 1, in step 114, the protocol designer uses the authoring tool to encode the eligibility criteria and the protocol schema for the clinical trial being designed. For the protocol schema, the authoring tool creates a graphical tool, called a knowledge acquisition (KA) tool (also considered herein to be part of the protocol authoring tool) that is used by protocol authors to enter the specific features of a clinical trial.

Fig. 9 is a flow chart detail of the step 114 (Fig. 1). In order to create an iCP, in a step 910, the protocol designer first selects the appropriate meta-model provided by the central authority in step 110. In most but not all cases, if the clinical trial protocol under development involves the testing of a particular treatment against a particular disease, then the step of selecting a meta-model involves merely the selection of the meta-model that has been created for the relevant disease category. In addition, in the embodiment described herein, each meta-model contains only a single list of relevant preliminary patient eligibility attributes and attribute choices. The step 910 of selecting a meta-model therefore also accomplishes a step of selecting one of a plurality of pre-existing lists of preliminary patient eligibility attributes. (Step 910A). As used herein, a list of eligibility attributes can be "defined" by a number of different methods, one of which is by "selecting" the list (or part of the list) from a plurality of previously defined lists of eligibility attributes. This is the method by which the list of preliminary patient eligibility attributes is defined in step 910A.

After the protocol author selects a meta-model, in step 912, the author then proceeds to design the protocol. The step 912 is a highly iterative process, and includes a step 912A of selecting values for the individual attributes in the preliminary patient eligibility attributes list; a step 912B of establishing further eligibility criteria for the protocol; and a step 912C of designing the workflow of the protocol. Generally the step 912A of selecting values for attributes in the preliminary patient attribute list will precede step 912B of establishing the further eligibility criteria, and both steps 912A and 912B will precede the step 912C of designing the workflow. However, at any time during the process, the protocol author might go back to a previous one of these steps to revise one or more of the eligibility criteria.

Fig. 10 is a flow chart of an advantageous method for the protocol author to establish the patient eligibility criteria. The protocol author is not required to follow the method of Fig. 10, but as will be seen, this method is particularly advantageous. The method of Fig. 10 is shown as a detail of step 914 (Fig. 9), which includes both the steps of selecting values for preliminary patient eligibility attributes and for establishing further eligibility criteria (steps 912A and 912B), rather than as being a detail of step 912A or 912B specifically, because the method of Fig. 10 can be used in either step above, or in both separately, or in both together.

The method of Fig. 10, sometimes referred to herein as an accrual simulation method for establishing patient eligibility criteria, substantially solves the problem mentioned above in which after finalizing a clinical trial protocol, engaging study sites and beginning the enrollment process, it is finally found that the eligibility criteria for the study are too restrictive and that with such criteria it is not possible to enroll sufficient patients in the trial. As mentioned above, these accrual delays are among the most costly and time consuming problems in clinical trials. The method of Fig. 10 addresses this problem by tapping an existing database of patient characteristics (database 116 in Fig. 1) as many times as necessary during the step 912 of designing the protocol, in order to choose eligibility criteria which are likely to enroll sufficient numbers of patients to make the study worthwhile. Generally the effort is to find ways to broaden some or all of the eligibility criteria just enough to satisfy that need, while maintaining



sufficient specificity in the study sample to ensure that the patients being treated are sufficiently similar in respect to clinical conditions, co-existing illnesses, and other characteristics which could modify their response to treatment.

Referring to Fig. 10, in step 1010, the protocol author first establishes initial patient eligibility criteria. Depending on which sub-step(s) of step 914 (Fig. 9) is currently being addressed, this could involve selecting values for the attributes in the previously selected patient eligibility attribute list, or establishing further eligibility criteria, or both. In step 1012, an accrual simulation tool runs the current patient eligibility criteria against the accrual simulation database 116 (Fig. 1), and returns the number or percentage of patients in the database who meet the specified criteria. If the database includes a field specifying each patient's location, then the authoring tool can also return an indication of which clinical sites are likely to be most fruitful in enrolling patients.

In one embodiment, the accrual simulation database includes one or more externally provided patient-anonymized electronic medical records databases. In another embodiment, it includes patient-anonymized data collected from various clinical sites which have participated in past studies. In the latter case the patient-anonymized data typically includes data collected by the site during either preliminary eligibility screening, further eligibility screening, or both. Preferably the database includes information about a large number of anonymous patients, including such information as the patient's current stage of several different diseases (including the possibility in each case that the patient does not have the disease); what type of prior chemotherapy the patient has undergone, if any; what type of prior radiation therapy the patient has undergone; whether the patient has undergone surgery; whether the patient has had prior hormonal therapy; metastases; and the presence of cancer in local lymph nodes. Not all fields will contain data for all patients. Preferably, the fields and values in the accrual simulation database 116 are defined according to the same CMT 112 used in the protocol meta-models and preliminary and further eligibility criteria. Such consistency of data greatly facilitates automation of the accrual simulation step 1012. Note that since the patients included in the accrual

simulation database may be different from and may not accurately represent the universe of patients from which the various clinical sites executing the study will draw, some statistical correction of the numbers returned by the accrual simulation tool may be required to more accurately predict accrual.

5           After accrual is simulated with the patient eligibility criteria established initially in step 1010, then in step 1014, the protocol author decides whether accrual under those conditions will be adequate for the purposes of the study. If not, then in step 1016, the protocol author revises the patient eligibility criteria, again either the values in the preliminary patient eligibility criteria list or in the further eligibility criteria or both, and loops back to try the accrual simulation step 1012 again. The process repeats iteratively until in step 1014 the protocol author is satisfied with the accrual rate, at which point the step of establishing patient eligibility criteria 914 is done (step 1018).

10           In an alternative implementation, the accrual simulation step 1012 is implemented not by querying a preexisting database, but rather by polling clinical sites with the then-current eligibility criteria. Such polling can take place electronically, such as via the Internet. Each site participating in the polling responds by completing a return form, either manually or by automatically querying a local database which indicates the number of patients that the site believes it can accrue who satisfy the indicated criteria. The completed forms are transmitted back to the authoring system, which then makes them available to the protocol author for review.

15           The authoring system makes them available either in raw form, or compiled by clinical site or by other grouping, or merely as a single total. The process then continues with the remainder of the flow chart of Fig. 10.

20           Returning to Fig. 9, both of the steps 912A and 912B preferably take advantage of concepts, terms and attributes already described in the CMT 112 (Fig. 1). The author may use a CMT browser for this purpose, which can either be built into the authoring tool, or a separate application from which the author may cut and paste into the authoring tool. In addition to the literal concept, terms and attributes entries, the CMT 112 preferable also contains "screen

25

questions", which are more descriptive than the actual entries names themselves, and which help both the protocol author and subsequent users of protocol to interpret each entry consistently.

The step 912C of designing the workflow, results in a graph like those shown in Figs. 3, 4, 6, 7 and 8 described above. As noted above, the authoring tool allows the protocol author to define not only patient management tasks, but also data management tasks. Such data management tasks can include such items as obtaining informed consent, completing forms regarding patient visits that have taken place, entering workflow progress data (e.g. confirmation that each patient management task identified for a particular visit was in fact performed; and which arm of a branch the patient has taken), and patient medical status information (e.g., patient assessment observations). In addition, preferably the concepts, terms and attributes used in the workflow graph make reference to entries in the CMT database 112. Even more preferably, as in the patient eligibility criteria, the authoring tool enforces reference to a CMT for all concepts, terms and attributes used in the workflow tasks. Again, a CMT browser may be used.

The result of step 912 is an iCP database, such as the one described above with respect to Figs. 2-8. As can be seen, the iCP contains both eligibility criteria and workflow tasks organized as a graph. The workflow tasks include both patient management tasks and data management tasks, and either type can be positioned on the graph for execution either pre- or post-enrollment.

In step 916, the iCP is written to an iCP database library 118 (Fig. 1), which can be maintained by the central authority. The iCP database library 118 is essentially a database of iCP databases, and includes a series of pointers to each of the individual iCP databases. In an embodiment, the iCP database library also includes appropriate entries to support access restrictions on the various iCP databases, so that access may be given to certain inquirers and not others.

For example, while certain clinical trial protocols are typically publicly available, such as those sponsored by public interest institutions, others, such as those funded by pharmaceutical companies, usually need to be maintained in strict confidence. Those which do not require

confidentiality might have no access restrictions other than subscription or membership in the central authority which maintains the iCP library 118, whereas those that do require confidentiality are accessible only by the sponsor itself and by specific clinical sites and SMOs who have signed confidentiality agreements with the sponsor. In an embodiment, the iCP database library 118 includes an access control list for each iCP database, which lists which users (including kinds of users) are allowed access to that database, and optionally what kind of access is permitted for each user (e.g., read/write access, or read access but not write access). The sponsor of each iCP is the only entity that can modify the access control list for the iCP.

Because the process of designing a clinical trial protocol can be extremely complex, usually requiring extensive medical and clinical knowledge, in one aspect of the invention the task is facilitated by allowing subprotocol components to be stored in a library after they are created, and re-used later in other protocols. Subprotocol components can themselves include subprotocol subcomponents which are themselves considered herein to be subprotocol components. In the object-oriented embodiments described above with respect to Figs. 2-8 and 11-25, the subprotocol components can be any object in an iCP, and subcomponents of such subprotocol components can be any sub-objects of such objects. Referring to Fig. 1, the subprotocol components are stored in a re-usable iCP component library 130, and they are drawn upon as needed by protocol designers in step 114, as well as written to by protocol designers (or sponsors) after an iCP or a portion of an iCP is complete.

In an embodiment, access to the subprotocol components in the re-usable iCP component library 130 is controlled in the same manner as is access to the iCP databases in iCP database library 118. This can be accomplished in various different ways in different embodiments. In one embodiment, the sponsor writes its subprotocol components into the library 130 with whatever granularity is desired. Each such component has associated therewith its own independent access control list, so access to the subprotocol components in the library 130 is controlled with the same granularity that the sponsor used in writing the components into the library 130. In another embodiment, access control lists are applied hierarchically within a subprotocol component,

permitting the granularity of access control to be finer than the granularity of objects written into the library 130 by the sponsor. In yet another embodiment the re-usable iCP component library 130 is physically the same as the iCP database library 118, and access control lists are associated with subprotocol components as well as with entire iCP databases.

5 In step 120, the central authority "distributes" the iCPs from the iCP database library 118 to clinical sites which are authorized to receive them. Authorization typically involves the site being part of the central authority's network of clinical sites, and also authorization by the sponsor of each study. In one embodiment, "distribution" involves merely making the appropriate iCP databases available to the appropriate clinical sites. In another embodiment,  
10 "distribution" involves downloading the appropriate iCP databases from the iCP database library 118, into a site-local database of authorized iCPs. In yet another embodiment, the entire library 118 is downloaded to all of the member clinical sites, but keys are provided to each site only for the protocols for which that site is authorized access. Preferably, the central authority maintains the iCP databases only on the central server and makes them available using a central application  
15 service provider (ASP) and thin-client model that supports multiple user devices including work stations, laptop computers and hand held devices. The availability of hand held devices allows the deployment of "intelligent" point of care data capture devices in which all protocol-specific, visit-specific and patient-specific required data elements, and their associated data validation rules, can be automatically created using information contained within the iCP. For example, an  
20 iCP can specify that if a patient exhibits evidence of an adverse event, additional data collection elements are required. Intelligent point of care data capture can detect the existence of an adverse event and add new required data elements to completely describe the event.

In step 122, the individual clinical sites conduct clinical trials in accordance with one or more iCPs. The clinical site uses either a single software tool or a collection of different software  
25 tools to perform a number of different functions in this process, all driven by the iCP database. In one embodiment, in which Protégé was used as a clinical trials protocol authoring tool, a related set of "middleware" components similar to the EON execution engine originally created

by Stanford University's Section on Medical Informatics, can be used to create appropriate user applications and tools which understand and which in a sense "execute" the iCP data structure. EON and its relationship to Protégé are described in the above-incorporated SMI Report Number SMI-1999-0801, and also in the following two publications, both incorporated by  
5 reference herein: Musen, et. al., "EON: A Component-Based Approach to Automation of Protocol-Directed Therapy, SMI Report No. SMI-96-0606, JAMIA 3:367-388 (1996); and Musen, "Domain Ontologies in Software Engineering: Use of Protege with the EON Architecture," Methods of Information in Medicine 37:540-550, SMI Report No. SMI-97-0657 (1998).

10 These middleware components support the development of domain-independent problem-solving methods (PSMs), which are domain-independent procedures that automate tasks to be solved. For example, the software which guides clinical trial procedures at the clinical site uses an eligibility-determination PSM to evaluate whether a particular patient is eligible for one or more protocols. The PSM is domain-independent, meaning that the same software  
15 component can be used for oncology trials or diabetes trials, and for any patient. All that changes between different trials is the protocol description, represented in the iCP. This approach is far more robust and scalable than creating a custom rule-based system for each trial, as was done in the prior art, since the same tested components can be reused over and again from trial to trial. In addition to the eligibility determination PSM, there is a therapy-planning PSM that directs  
20 therapy based on the protocol and patient data, and the accrual simulation PSM described elsewhere herein, among others.

Because of the ability to support domain-independent PSMs, the iCPs of the embodiments described herein enable automation of the entire trials process from protocol authoring to database lock. For example, the iCP is used to create multiple trial management  
25 tools, including electronic case report forms, data validation logic, trial performance metrics, patient diaries and document management reports. The iCP data structures can be used by multiple tools to ensure that the tool performs in strict compliance with the clinical protocol

requirements. For example, the accrual simulation tool described above with respect to Fig. 10 is implemented as a domain-independent PSM. Similarly, an embodiment can also include a PSM that clinical sites can use to simulate their own accrual in advance of signing on to perform a given clinical trial. A single PSM is used to simulate accrual into a variety of studies, because the patient eligibility criteria are all identified in a predetermined format in the iCP for each study. Another PSM helps clinical sites identify likely patients for a given clinical trial. Yet another PSM guides clinicians through the visit-specific workflow tasks for each given patient as required by the protocol. The behavior of all these tools is guaranteed to be consistent with the protocol even as it evolves and changes because they all use the same iCP. The tools can also be incorporated into a library that can be re-used for the next relevant trial, thus permitting knowledge to be transferred across trials rather than being re-invented each time.

Fig. 26 is a flow chart detail of step 122 (Fig. 1). The steps in Fig. 1 typically use or contribute to a site-private patient information database 2610, which contains a number of different kinds of patient information. Because this information is maintained in conjunction with the identity of the patient, these databases 2610 are typically confidential to the clinical site or SMO, and not made available to anyone else, including study sponsors and the central authority. In one embodiment, the patient information database 2610 is located physically at the clinical site. In another embodiment, storage of the database 2610 is provided by the central authority as a service to clinical sites. In the latter embodiment, cryptographic or other security measures may be taken to ensure that no entity but the individual clinical site can view any confidential patient information.

As shown in Fig. 1, the central authority also maintains its own "operational" database 124, containing patient-anonymized patient information. The operational database 124 can be separate from the confidential patient information database(s) 2610 on which case a patient anonymized version of the patient information database 2610, or at least portions of database 2610, are transferred periodically for inclusion in an operational database 124 (Fig. 1).

Alternatively, the two databases can be integrated together into one, with the central authority being denied access to sensitive patient-confidential information cryptographically.

Referring to Fig. 26, when a particular site is considering signing on to a clinical study for which it is authorized, it can first perform an accrual simulation, based on the data in its own patient information database 2610, to determine whether it is likely to accrue sufficient numbers of patients to make its participation in the study worthwhile (Step 2612). As mentioned, step 2612 is performed by a PSM which references the preliminary eligibility criteria and, in some embodiments, the further eligibility criteria for the candidate study. This PSM queries the site's information database 2610. Preferably the patient information database 2610 includes an indication of which patients are already enrolled in clinical trials, and excludes them from the predicted accrual count reported by the local accrual simulation tool.

After the clinical site has decided to proceed with a study, then it can use either a "Find-Me Patients" tool (step 2614) or a "QuickScreen" tool (step 2616) to identify enrollment candidates. The "Find-Me Patients" tool is either the same or different from the local accrual simulation tool, and it operates to develop a list of patients from its patient information database 2610 who are likely to satisfy the eligibility criteria for a particular protocol. Again, this local "Find-Me Patients" tool makes appropriate queries to the patient information database 2610 for patients who are believed to satisfy the preliminary eligibility criteria for the subject protocol.

The QuickScreen tool, on the other hand, for each candidate patient, compares that patient's characteristics with the preliminary eligibility criteria for all of the studies which are relevant to that clinical site. In one embodiment, this is an automated process driven by the eligibility criteria for all the relevant protocols, and which queries the site's patient information database 2610. In an embodiment, the QuickScreen tool references only those patient characteristics already stored in the patient information database 2610, for which informed consent from the patient was not originally required, or for which the patient previously gave his or her informed consent for a different purpose. Note that the iCPs in the present embodiment do not actually include the preliminary criteria with the iCP. Rather, the preliminary criteria are



provided separately by the central authority in a unified "QuickScreen database". In another embodiment, the preliminary eligibility criteria can be either included directly in the iCP or identified by pointer from the iCP. As used herein, a database "identifies" certain information if it contains the information or if it points to the information, whether or not through one or more levels of indirection.

The studies that the site includes in the QuickScreen step 2616 for a given candidate patient might be all studies for which the site has authorization. Alternatively, the list of candidate studies can be limited to a particular disease area, and/or by the patient's own stated preferences, and/or by other study selection criteria applied by the site. If the QuickScreen step 2616 involves manual entry of newly obtained patient data, then preferably such data is added to the patient information database 2610.

If the candidate patient is determined to satisfy the preliminary eligibility criteria for one or more clinical trials, in step 2616, then in step 2618, the clinical site evaluates the candidate patient's medical characteristics against the further eligibility criteria for one or more of the surviving studies. This step can be performed either serially, ruling out each study before evaluating the patient against the further eligibility criteria of the next study, or partially or entirely in parallel. Preferably the step 2618 for each given study is managed by the workflow management PSM, making reference to the iCP for the given study. The iCP may direct certain patient assessment tasks which are relevant to the further eligibility criteria of the particular study. It also directs the data management tasks which are appropriate so that clinical site personnel enter the patient assessment results into the system for comparison against the further eligibility criteria. Furthermore, as described in more detail elsewhere herein, the iCP can further direct the obtaining of informed consent either at the beginning of the further eligibility evaluation step 2618, or at an appropriate time in the middle when informed consent is required before proceeding. Moreover, where possible, all data entered into the system during step 2618 is recorded in the clinical site's patient information database 2610.

After step 2618, if the patient is still eligible for one or more clinical trials, then in step 2620, the workflow management tool directs and manages the process of enrolling the patient in one of the trials. The fact of enrollment is recorded in the patient information database 2610. In step 2622, the workflow management tool, governed by the iCP database, directs all of the workflow task required at each patient visit in order to ensure compliance with the protocol. As mentioned, in accordance with the protocol, information about the patient's progress through the workflow tasks is written into the patient information database 2610, as is certain additional data called for in the data management tasks of the protocol. In one embodiment, the workflow management tool records performance/non-performance of tasks on a per patient, per visit basis. An another embodiment, more detailed patient progress information is recorded.

Returning to Fig. 1, as can be seen, patient-anonymized medical information as well as workflow progress information is uploaded from the patient information databases 2610 at each of the clinical sites in the network, to a central operational database 124. In various embodiments, some or all of this data is uploaded immediately as created, and/or on a periodic basis. The clinical study sponsors have access to the data in order to permit real time or near-real-time (depending on upload frequency) monitoring of the progress of their studies (Step 126), and the central authority also analyzes the data in the operational database 124 in order to rate the performance of each site against clinical site performance metrics (Step 128).

Such performance metrics include a site's accrual performance (actual vs. expected accrual rates), and the site's ability to deliver timely, accurate information as trials progress. The latter metrics can include such measurements as the time to complete tasks, the time from visit to entered CRF, the time from visit to closed CRF, the time from last visit to closed patient, and the time from last patient last visit to closed study. Prior art systems exist for collecting site performance data, but these systems have captured only very narrow metrics such as completion of case report forms, and the number of audits that have been conducted on the site. The prior art systems are also entirely paper-based. Most importantly, the prior art systems evaluate site performance only for a single specific study; they do not accumulate performance metrics across

multiple studies at a given clinical site. In the embodiment described herein, however, the central authority gathers performance data electronically over the course of more than one study being conducted at each participating clinical site. In step 128 the central authority evaluates each site's performance against performance metrics, and these evaluations are based on each site's proven and documented past performance, typically over multiple studies conducted. Preferably, the central authority makes its site performance evaluations available to sponsors such that the best sites can be chosen for conducting clinical trials.

Study sponsors also have access to the data in the operational database 124 in order to identify promising clinical sites at which a particular new study might be conducted. For this purpose, the patient information that has been uploaded to the operational database 124 includes an indication of the clinical site at which the data was collected. The sponsor then executes a "Find-Me-Sites" PSM which queries the operational database 124 in accordance with the iCP or preliminary eligibility criteria applicable to the new protocol, and the PSM returns the number or percentage of patients in the database from each site who satisfy or might satisfy the eligibility criteria.

As used herein, a given event or value is "responsive" to a predecessor event or value if the predecessor event or value influenced the given event or value. If there is an intervening step or time period, the given event or value can still be "responsive" to the predecessor event or value. If the intervening step combines more than one event or value, the output of the step is considered "responsive" to *each* of the event or value inputs. If the given event or value is the same as the predecessor event or value, this is merely a degenerate case in which the given event or value is still considered to be "responsive" to the predecessor event or value. "Dependency" of a given event or value upon another event or value is defined similarly.

The foregoing description of preferred embodiments of the present invention has been provided for the purposes of illustration and description. It is not intended to be exhaustive or to limit the invention to the precise forms disclosed. Obviously, many modifications and variations will be apparent to practitioners skilled in this art. In particular, and without

